AFCRL 67-0313

AD657358

1 MAY 1967

CLIMATIC NORMALS AS PREDICTORS

Part I: Background

by

ARNOLD COURT

PROFESSOR of CLIMATOLOGY

SAN FERNANDO VALLEY STATE COLLEGE

SCIENTIFIC REPORT NO. 1

CONTRACT AF19 (628)-~~6176~~ 5716

PROJECT NO. 8624

TASK NO. 862402

REPORT NO. ~~67-02-1~~ 82-67-1

SAN FERNANDO VALLEY STATE COLLEGE FOUNDATION

NORTHRIDGE

(Los Angeles)

Calif. 91324

CONTRACT MONITOR: IVER A. LUND, CREW

prepared for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES

OFFICE OF AEROSPACE RESEARCH

UNITED STATES AIR FORCE

BEDFORD, MASSACHUSETTS

01730

40

AFCRL 67-0313

CLIMATIC NORMALS AS PREDICTORS

PART I: BACKGROUND

BY

ARNOLD COURT
Professor of Climatology
San Fernando Valley State College
Northridge (Los Angeles) Calif., 91324

Scientific Report No. 1

Contract AF 19(628)-5716

Project No. 8624 Task No. 862402

1 May 1967

Contract Monitor: Mr. Iver A. Lund, C R E W

Distribution of this document is unlimited.

Prepared for

Air Force Cambridge Research Laboratories

Office of Aerospace Research

United States Air Force

Bedford, Massachusetts

01730

ABSTRACT

Preliminary to investigating the length of the climatic period whose average gives the best (minimum variance) estimate of the next year's value, previous studies are examined and the results of five are replotted onto a standard scale. All indicate that prediction one year ahead from an average based on only 20 years, or so, is better than one from a standard "climatic normal" of 30 years. Monte Carlo simulation of the prediction process suggests that slight changes with time in the means, whether real or caused by instrumental or observational changes, in most climatic records reduce the record length for optimum prediction.

CLIMATIC NORMALS AS PREDICTORS

Arnold Court

1. BACKGROUND

1.1 Normals

Climatic normals are averages of the values of a climatic element during many years. They describe the climate of a place or region, specifically during the period for which they are computed. Often, however, this description is extrapolated to estimate future climatic conditions. This report is concerned with such predictive use of normals, and specifically with determining the number of years whose average, or other statistic, offers the best estimate of conditions one to ten years later.

The basic question which led to the present study concerned the proper procedure for summarizing climatic data from a new location, or by a new method of measurement. Should each year's observations be added to the preceding ones, or should a moving average be used, dropping out the earliest year so that the averages are based on a constant number of years; if so, what should that number be? The immediate problem involved upper air observations, specifically wind observations, whose accuracy, precision, and completeness increase each year. Since the concept of climatic normals is not generally applied to upper air data, a review of the entire question was advisable.

Originally, more than a century ago, climatic normals were considered to approximate the "true" climate which, like the everlasting hills

and the plants, was assumed to have been constant since the Deluge. Under this assumption of a stable climate, subject only to random variations from year to year but essentially invariant over the centuries, the longest record gave the best average. For more than a century, the standard error of a mean of k independent observations has been known to be

$$s_{\bar{x}} = \sigma / \sqrt{k} \quad , \qquad (1)$$

where σ is the standard deviation of the individual observations about their true mean. A 100-year mean is twice as precise as one based on only 25 years, as an estimate of the "true" climatic mean. Many people have computed standard deviations of climatic series to determine the number of years, k , needed to give a standard error less than some arbitrary value, such as 1 degree for temperature or 0.1 inch or 0.5 cm for precipitation.

However, during the present century the concept of an invariant climate has been replaced, gradually, by the realization that climate fluctuates over the decades, centuries, and millenia. Whether these variations have any regularity is hotly debated; proponents have been unable to muster statistical proof of the reality or importance of postulated solar and lunar cycles, sub-cycles, and super-cycles. But certainly conditions have been warmer or drier during certain spans of years than during preceding or following spans of equal length.

Such fluctuations in short-period means cause estimates of the standard deviation of a climatic element to increase with the length of the record. A similar increase arises from inhomogeneities in the observations, caused by changes in instrumentation, exposure, and method of observation. The total increase in σ is greater then

$\sqrt{k/(k-1)}$, the correction for the size of a random sample. It reduces, by an unknown amount, the apparent increase in precision shown by Eq. (1).

Furthermore, Eq. (1) is based on the assumption of random sampling from statistically independent observations, and the existence of fluctuations suggests, but does not prove, that these conditions may not be met. Fluctuations, although clearly evident in the record of a climatic element, may still have been the result of a random process. A true coin can fall heads several times in succession, and the longer the tossing continues, the greater the chance of an arbitrarily long run. Some climatic fluctuations, or sequences of fluctuations, however, appear to exceed significantly what would be expected in a random series, suggesting that k successive values of the element are not independent, and Eq. (1) may not apply strictly.

The Working Group on Climatic Fluctuations aptly asked, in its comprehensive report to the World Meteorological Organization's Commission for Climatology at Stockholm in August, 1965: "If non-randomness is present, does it take the form of persistence, trend, periodic fluctuations, aperiodic fluctuations, or perhaps some combination of these?" To answer this question, the Working Group, under the able chairmanship of Dr. Murray Mitchell, suggested a series of elaborate statistical tests of a long and homogeneous record.

Without such information, climatic normals must be evaluated empirically, by how well they describe the climate, or predict future values. Efforts at defining descriptive normals are summarized in the next Section, and predictive ones in later Sections. Other descriptive statistics, primarily order statistics such as the median, the quartiles or other fractiles, will be discussed in a subsequent Chapter, to be issued later.

1.2 <u>Descriptive normals</u>

The validity and utility of a climatic average depend on the homogeneity of the original observations, as well as on any natural fluctuations during the period of observation. Few climatic records have been obtained by constant procedures from the same (or equivalent) instruments in unchanging exposures for more than a few decades. To reduce the effects of these two factors, in 1935 the International Meteorological Organization broke the tradition of using the entire "period of record" for climatic normals, and adopted instead the 30-year period 1901-1930.

This recommendation has been followed by most weather services, and was reiterated by the successor World Meteorological Organization in 1957. It adopted this definition:

> <u>Climatological standard normals</u>: Averages of climatological data computed for consecutive periods of 30 years, the first of which started on 1 January 1901.

Averages for any other set of "at least three consecutive 10-year periods" are called <u>normals</u>, and averages "for any period of at least ten years starting on 1 January of a year ending with the digit 1" are <u>period averages</u>. (Mitchell's Working Group suggested changing these definitions to start the periods in 1900, etc., for greater ease in punch-card sorting.)

Since the adoption of the 30-year climatological standard normal, many studies have been made of its "representativeness" and other properties. Some compared normals for successive 30-year periods, such as 1871-1900 and 1901-1930, or for overlapping periods, such as 1901-1930 and 1911-1940. Others have compared 30-year normals with means for much

longer periods, a century or more. Many of these studies found significant differences between the various normals and means, and therefore questioned the utility or value of a "climatological standard normal" or other fixed-period normals.

More broadly, other investigators have examined how closely the mean of k successive observations, not restricted to $k = 30$, approaches the mean of a longer period that includes those k years. Carruthers [1945] tabulated departures of means of varying length, from 1 to 70 years, from the 202-year (1742-1943) mean rainfall of Great Britain, and concluded that 35 years offered an adequate compromise between precision and available observations. This validated long-standing British custom of using a 35-year mean, so British rainfall normals are still computed for 35 years, rather than the 30 urged by WMO.

Lenhard and Baum [1954] used confidence limits based on random sampling from a normal population to determine the smallest value of k yielding a mean monthly temperature that would "describe the record;" tests showed both normality and independence were acceptable for January and July temperatures at most of their seven stations. The minimum number of years for which the mean had a standard error of less than 1 deg. F. varied from 10 at San Diego to 73 at Bismark in January, and was generally smaller in July. Coffin [1954] offered a regression technique for estimating, from data for 10 or even fewer years, the "normal most representative of the present rainfall regime" at stations in Washington and Oregon.

1.3 Predictive averages

Describing the past climate, however, is not the primary use to which climatic normals are put. Description of conditions during a specified period is of value only for the study of climatic variations from period to period, and for comparisons either with similar normals from other places or with other phenomena, such as plant distribution or disease incidence. Even such comparisons, among which climatic classification is a common example, are largely for extrapolation: they are assumed, implicitly, to apply beyond as well as during the normal period.

In most applications, normals or other averages are used to <u>predict</u> future conditions. The predominant use of a climatic normal is to estimate the crops that can be grown, the heat that will be required, the water that will be available, the clothing that will be worn, the transportation interruptions that will occur, or any of the manifold likely effects of weather on man's diverse activities and interests. Such applications may be for next year only, or for the next decade, or even for the entire next century, in the case of large water projects.

Only in the past dozen years have climatic averages been examined specifically with respect to this primary application of predicting future values. Lenhard and Baum [1954] recognized that "extrapolation of the record [while] of dubious validity ... is implicit in almost all activities." Use of "a normal temperature" for design, they said, assumes "that the normal temperature used will be characteristic of times to come. Probably the best forecast that can be made is to use the normal from the part of the record nearest to the point of extrapolation." But they did not examine, specifically, the predictive

accuracy of climatic averages or normals.

How well a k-year average of a given climatic element will predict the value of that element m years later was studied by Brier [1955], Beaumont [1957], Enger [1959a], Craddock and Grimmer [1960], and Drozdov et al. [1965]. It is being investigated, in even greater detail, in the present study. Brier's work is not published, Beaumont's formula contains annoying misprints, Enger presented his procedure piecemeal, as did Craddock and Grimmer, and Drozdov et al. used a somewhat different approach. Hence all their findings will be presented in Figs. 1 through 7 and Table A, and discussed in Sections 1.5 and 1.6, using the notation adopted for the present study.

In a time series of n climatic observations, three intervals are involved in the discussion: the number, k , of antecedent observations from which a mean (or other statistic) is computed, the length, ℓ , of the future period for which the mean (or other statistic) is to be estimated from that for the k years, and the "lag," m , between the two periods:

```
      ------k------  ___m___  --ℓ--
• • • • • • • • • • • • • • • • • • • • • • • • • • • • • • • • • •
1   5     10     15     20     25     30       n
```

The extrapolation variance, $S^2_{k\ell m}$, in using a k-year average to estimate an ℓ-year average beginning m years later is computed, from a series of n observations x_i ordered in time from x_1 to x_n, as

$$S^2_{k\ell m} = \frac{1}{n-k-\ell-m+2} \sum_{i=1}^{n-k-\ell-m+2} \left[\frac{1}{\ell} \sum_{j=m}^{m+\ell-1} x_{i+j+k-1} - \frac{1}{k} \sum_{j=0}^{k-1} x_{i+j} \right]^2 \qquad (2)$$

Its square root, $S_{k\ell m}$ is the standard error of extrapolation. The mean prediction error, $Q_{k\ell m}$, is obtained by taking the absolute value, rather than the square, of the difference.

In all previous studies except that of Drozdov et al., prediction was of single values, rather than ℓ-year means, so that for $\ell = 1$ the first sum inside the parentheses is simply x_{i+k}. In all studies, except part of Enger's, only next year's value, or the k + 1st, was estimated, so that $m = 1$, further simplifying the expressions. When both ℓ and m are unity, $S_{k11} = S_k$.

All investigators were interested primarily in finding the value of k for which $S^2_{k\ell m}$ (or $S_{k\ell m}$) or $Q_{k\ell m}$ was smallest. This minimizing of k will here be denoted as k* ("k-star"). Implicitly, in the previous studies k* was assumed to indicate the "optimum length of record" for prediction. To determine k*, Beaumont used S^2_k, the "average error mean square" or AEMS, which he denoted by σ. Brier used "total square error" and Enger both "mean square error" and "root mean square error," the latter denoted as E_k. The "residual variance" of Craddock and Grimmer is equivalent to S^2_k. Only Drozdov et al. used the mean absolute difference $Q_{k\ell m}$.

1.4 Significance

Few estimates have been offered of the statistical significance of the minimum values of S^2_k or S_k in these previous investigations. In discussion of Beaumont's [1957] paper, van Hylckama [1958] reported that

two sets of 51 random numbers treated by Beaumont's procedure, essentially Eq. (2), gave minimum values of AEMS (or S_k^2) at $k^* = 5$ and $k^* = 30$. Citing various findings that "yearly precipitation data show rarely, if ever, any persistency or definite periodicity...", he concluded that "the AEMS is a random number taken from random data and cannot possibly be indicative of future events." Beaumont agreed that random numbers give graphs of S_k vs. k similar to his results for annual streamflow and precipitation, but insisted that they showed less consistency than his data.

Enger [1959a] analyzed as though they were maximum temperatures "a large number of sets of deviates" obtained from random numbers. He presented four examples of random sample results (two for $m = 1$ and two for $m = 1,5,9$ combined) and four graphs of "total square error." These total curves, which apparently indicate the total square difference for 10 samples each, all decrease monotonically to $k^* = 30$, the largest value used. Enger then estimated, from these results, the probability of the various $\min S_k^2$ in his data.

In random sampling from a normal population with mean 0 and variance σ^2, a k-element mean, $\bar{x}_k$, is normally distributed with mean 0 and variance σ^2/k. The difference of two independent normal variables also has a normal distribution, with mean equal to the difference of their means and variance equal to the sum of their variances. Hence the difference between one observation and a k-element mean, not including it, is normally distributed with mean 0 and variance $\sigma^2 + \sigma^2/k$.

Thus, in random sampling from a normal population, the mean square difference between a k-year mean and the population mean decreases as

$1/k$, Eq. (1), while the mean square difference between a k-year mean and a random single value (not included in the k years) decreases as $1 + 1/k$. The standard error of extrapolation, S_k , is $\sqrt{k+1}$ times the standard error of the mean:

$$S_k = s_{\bar{x}} \sqrt{k+1} = \sigma \sqrt{1 + 1/k} \qquad (3)$$

These relations offer a first approach to assessing the significance of the behavior of S_k^2 with increasing k . If S_k^2 decreases more rapidly, for some range of k , than $1 + 1/k$, a minimum may represent more than van Hylckama's "random number taken from random data." To provide a visual indication of the behavior of S_k^2 compared to $(1 + 1/k)\ \sigma^2$, curves of the theoretical ralation have been added to graphs, discussed in Section 1.5, showing the results of the previous studies. To draw such curves, some estimate of σ^2 is needed; arbitrarily, σ^2 was taken as equal to 1,2,3,··· or 10, 20, 30, ··· at $k = 50$, so that the curves become horizontal at $k = 50$. The difference at $k = 30$, however, is so slight that the curves offer general guidelines for all computations of S_k^2 .

Since the "true mean" is assumed to be constant, this relation should not depend on m , the separation between the k years and the year for which the estimate is made. But it does depend on ℓ , the length of the period to which the forecast is applied, if $\ell > 1$. The difference between a k-element mean and an ℓ-element mean is normally distributed, under the previous assumptions, with mean 0 and variance $(1/k + 1/\ell)\ \sigma^2$. When a k-year mean is used as a predictor of the mean of another k-year period, the extrapolation variance is $2\ \sigma^2 / k$, twice the variance of

either mean with respect to the "true" climatic value.

These relations are based on the assumption that the "true mean" is constant, and that individual observations all have the same variance about it. That is, they assume that the sequence of observations forms a stochastic time series that is stationary in both first order (for mean) and second order (for variance). Actual climatic data, however, may violate one or both of these assumptions. Means may be changing with time, either slowly or by jumps, and so may variances. These changes may be either natural, true climatic changes, or they may be observational, arising from changes in the instruments used and their exposures, as well as in the manner of reading and the procedures for summarizing those readings. Consequences of such non-stationarity will be discussed in a later Section.

1.5 Previous Studies

Results were presented in the previous studies almost entirely as graphs of S_k^2 or S_k, or of "total square error", as a function of k; usually results for each station and climatic element were graphed separately, often on differing scales. All these results have been scaled from the graphs and replotted on a common basic diagram, on which lines of $(1 + 1/k)\sigma^2$ have been added. Some details of the various studies are summarized in Table A.

Only annual values, of stream flow at three places and precipitation at ten, all in Western United States, were studied by Beaumont (Fig. 1). His one-year predictions ($m = 1$) used means of $k = 5[5]35$ preceding years, with n "of sufficient length to provide 30 and 35 year means."

Table A. Previous estimates of minimum value of standard error of extrapolation, $S_{k,m}$, using k-year mean to estimate value m years later, based on data from indicated number of stations over n years. (Last column refers to Figure in which data are given).

Element	Period	Stns.	n	k	m	min $S_{k,m}$	Author	Fig.
YEARLY								
river flow	year	3	35+	5[5]35	1	15-25	Beaumont	1
tot. prec.	"	10	"	"	"	10-35	"	
random	--	2	51	"	1	5,30	van Hycklma	-
mean temp.	year	22a	45-203	1[1]50	1	31-35c	Craddock	2
" "	"	57a	40-160	"	"	11-15c	& Grimmer	-
MONTHLY								
mean temp.	Jan	4	60+	5[1]40	1	8,16,17,20	Brier	3
tot. prec.	"	"	"	"	"	6,15,15,19	"	
mean temp.	Jan	11	50?	1[1]30,35	"	15-28	Enger	4
" "	Jul	"	"	"	"	2-35	"	5
DAILY								
max. temp.	day*	10	21	3,7,10[5]30	1	7-30	Enger	6
" "	"	"	"	"	1,5,9	10-30	"	7
random	#	many	?	"	1	5-30	"	

* predictor is average of 31 days centered on predictand day
\# predictor is average of 5 numbers
a: 22 "homogeneous" stations, 57 "nonhomogeneous" stations
c: 5-year range shown as most frequent

For the three rivers, minimum values of S_k are attained at $k* = 15$, 20, and 25 years, but only the 20-year minimum, for the Columbia River, was particularly sharp; a composite graph of all three (not reproduced here) "indicated that a 15-year mean is the best." The graphs of the 10 annual precipitation records show minimum values of S_k from $k* = 10$ to $k* = 35$ years. A weighted composite graph (not reproduced here) showed the minimum at $k* = 20$ years. (Objections, advanced by van Hycklyma [1958], that similar results could be obtained from random numbers, were discussed in Section 1.4.)

In the only other investigation of annual values, Craddock and Grimmer (1961) found that annual temperatures, one year ahead, can be predicted adequately from means for the preceding 10 to 30 years of homogeneous observations, "but shorter averages are definitely preferable for use with non-homogeneous records." Without referring to previous American work, they tried both unweighted and exponentially weighted k-year means from 79 long-record stations, of which only 22 were adjudged homogeneous. Exponentially-damped values generally gave larger values of S_k, so their study concentrated on k-year means, with k going by 1-year steps from 1 to 50, or at least to $n - 15$, where n is the total number of years of record.

Their results for six stations are replotted in Fig. 2, in terms of S_k^2 rather than S_k. "In all except Bermuda," they remarked, "the values of S_k decrease quite sharply as k increases from 1 to about 5," generally in accord with the theoretical $1 + 1/k$ relation, which they did not use. "As k continues to increase there is a zone in which S_k changes very little, but with still larger values of k, S_k shows a definite and

unmistakable increase. These features, the rapid decrease to start with, the flat zone, and the terminal increase, are present in all but 5 of the 85 records. The Bermuda record is a curiosity, not repeated at any other station, in which a better prediction is obtained by last year's values than by averages over any period of past years."

Mean monthly values, not annual, of temperature and precipitation in January and July at St. Louis, Boston, Greenwich, and Copenhagen, were studied by Brier in an unpublished 1954 study, apparently unknown to Beaumont. From tabulations of total square error for $k = 5[1]40$, he concluded "that little is to be gained by using more than 15 to 20 years," according to Enger (1959a). Brier's results, replotted from Enger's diagram, are presented in Fig. 3; all values were divided by 60, the approximate number of years used, to express them as S_k^2. In addition, the Copenhagen values have been converted from metric to English units, for ready comparison with those of other studies, and the Copenhagen precipitation variances multiplied by 1,000 to correct an apparent error: Enger's graph indicates a total square error of about 40 mm^2 which would be a mean square error of only .001 in^2. Likewise, the extrapolation variance for Copenhagen temperatures seems too small by a factor of two or more.

Values of S_k^2 from Brier's data (Fig. 3) generally follow the theoretical $1 + 1/k$ curves, at least as far as $k = 20$ or so. But the extrapolation variances based on longer periods tend to increase, so that 30-year normals give definitely larger extrapolation variances than those for shorter periods.

Enger continued Brier's work for his master's thesis, completed a year before Beaumont's paper appeared but not published, in summary, until

two years later. He used mean monthly temperatures for January and July at 11 U. S. "climatological benchmark" stations for $k = 1[1]30$ or 35, and also maximum daily temperatures on four dates in January and July at 10 other U. S. stations, but only for $k = 3,7,10[5]30$, rather than for every value of k. Both sets of temperatures were tabulated for prediction one year ahead ($m = 1$). In addition the k-year means of maximum daily temperatures were compared to values $m = 1, 5$ and 9 years later, as discussed in the next Section.

Enger concluded Brier's, Beaumont's, and his own investigations of "annual, monthly, and daily climatological variables all agree that a relatively short period of record, of the order of 15 to 25 years, is best for estimating values one year ahead.... Very short periods of record are adequate for climatological prediction purposes," and "extending a climatological record backwards" may not be worth the effort, at least for predictive purposes.

Enger's results have been scaled from the individual diagrams, with the "root-mean-square of errors of prediction" squared and the "total square error" divided by sample size, and replotted on Figs. 4, 5, 6, and 7. On the first two, for mean January and July temperatures, respectively, at the 11 "benchmark" stations, the curves follow the theoretical $1 + 1/k$ relation fairly well, especially in January. The extrapolation variance in January is about four times that in July, as indicated by the change in ordinate by a factor of four; at Dickinson, N.D., the January variance of around 100 (deg F)^2, requiring a special scale, is an order of magnitude greater than the July value.

Thus four different studies show that next year's annual or monthly temperature or precipitation can be estimated with smaller variance from the mean of the preceding 15 to 25 years than from a 30-year mean. These findings are generally corroborated by the present investigation, to be presented in a later report.

1.6 Prediction beyond next year

Only two previous studies have been concerned with using k-year climatic averages to predict, more than one year ahead, either an ℓ-year mean or a single value m years ahead. Part of Enger's (1959a) study concerned prediction of the maximum temperature on a specific date, not only next year but also five and nine years later. In a related paper, Enger (1959b) determined that the maximum temperature on a specific date could be predicted with smaller variance from the mean maximum temperature, during k preceding years, of the 31-day period including that date than from the maximum temperatures on the only same date in k preceding years, or from means over less than 31 days.

The average extrapolation variance for maximum temperatures on the 15th, 20th, 25th, and 30th days of January and July, at five U.S. stations, using date-centered 31-day mean maximum temperatures for the preceding k years, is shown on Fig. 6. The corresponding mean extrapolation variance, further averaged for prediction $m = 1, 5$, and 9 years ahead, is shown in Fig. 7.

The curves are much more regular than those for monthly temperatures, largely because k was taken at 5-year increments, rather than 1-year. But many show no significant decrease in S_k^2 from $k = 3$ to $k = 7$, and few

show any increase as k grows larger. More significant, at half the stations (Blue Hill, Tucson, Tatoosh, and San Diego in January, and Key West in July) the mean extrapolation variance for prediction 1, 5, and 9 years ahead is smaller than for prediction only 1 year ahead. Enger did not comment on this surprising feature. Since the 1-year-ahead values are included in the mean variance, prediction 5 and 9 years ahead, at these stations, must have substantially smaller variance than prediction only one year ahead.

Rather than a single value m years ahead, the mean value over the next $\ell = 5$, 10, or 15 years was estimated from k-year antecedent means by Drozdov et al. [1965]. But they used only $k = 10$, 25, and 50 years, which other studies suggest is too gross to reveal minimum values of $S_{k,\ell,m}$ and presumably also of their criterion, the mean absolute difference, $Q_{k,\ell,1}$. Table 1 of Drozdov et al. shows that in general 50-year means gave smaller mean differences (Q) from subsequent 10-year means than did either 10 or 25-year means. Also, 50-year means differed slightly less from the means of subsequent 5, 10, and 15-year periods than did either 25 or 10-year means.

Similarly, Davitaia [1966] found that 10-year means were predicted with smaller errors from the means of larger antecedent periods than from shorter ones, but examined only $k = 10$, 30, 50, and 100 years.

1.7 Random numbers

The general tendency, shown in the studies of previous investigators, for S_k^2 to reach a minimum in less than 30 years, rather than to decrease according to $1 + 1/k$, suggests that the climatic series are not made up of random samples from a homogeneous population. As already discussed

Table 1. Number of cases, by months, in which increasing the period of averaging decreased (-), increased (+), or did not change (0) the difference between that average and that for the following ten years.

Month	50 years vs. 10 years			50 years vs. 25 years		
	-	+	0	-	+	0
JAN	8	1	0	6	1	2
FEB	8	.0	1	6	1	2
APR	4	4	1	1	5	3
JUL	4	2	3	2	3	4
OCT	5	2	2	5	4	0
J+F+O	29	9	7	20	14	11
TOTAL	21	3	3	17	6	4

Table 2. Average difference between mean temperature for 10, 25, and 50-year periods and mean temperatures of following 5, 10 and 15 year-periods.

Month	for next 5 yrs			for next 10 yrs			for next 15 yrs		
	10	25	50	10	25	50	10	25	50
JAN+FEB	1,4	1,1	1,1	1,2	1,1	1,1	1,2	1,0	1,0
MARCH	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4
JULY	1,3	1,3	1,1	1,3	1,3	1,2	1,3	1,3	1,2
OCT	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3
TOTAL	2,1	2,1	2,0	2,1	2,0	2,0	2,1	2,0	2,0

Tables 1 and 2 of Drozdov et al.

(Sec. 1.4), slow changes with time in means or in variances could cause the observed behavior of S_k^2 . To investigate the effects of such changes, a Monte Carlo approach was adopted.

Random normal numbers (mean zero, variance unity) were used to compute S_k^2 according to Eq. (2), for two samples of $n = 100$ and one of $n = 1,000$, called Samples 1, 2, and 3, respectively. Samples 1 and 2 both showed $k^* = 16$, but Sample 3 showed an almost monotonic decrease of S_k^2 to $k^* = 50$, the limit of the calculation.

Then each sample was biased in mean and variance, separately. Each sample was divided into thirds -- of 33, 34, and 33 (or 333, 334, 333) "years." For test 1, each random number in the first third was decreased by 0.5, each number in the final third was increased by 0.5. For Test 2, the increments were -1.0, 0.0, and +1.0, to bias the means more strongly. For Test 5, the numbers in the first third were multiplied by 1.5, those in the middle third were unchanged, and those in the final third multiplied by 0.75; for Test 6, the factors were 2.0, 1.0, and 0.5, to bias the variances more strongly.

Each sample was then divided into fifths, of 20 (or 200) numbers each, and the following biases applied:

Fifth:	1st	2nd	3rd	4th	5th
Test 3	-0.6	-0.2	0.0	+0.2	+0.6
Test 4	-2.0	-1.0	0.0	+1.0	+2.0
Test 7	X 1.6	X 1.3	X 1.0	X 0.7	X 0.4
Test 8	X 3.0	X 2.0	X 1.0	X 0.5	X 0.25

Results of these nine tests applied to each sample are shown on two figures, one for the biases of the means, the other for the biases of the

variances; on each, a line labelled "0" shows the variation of S_k^2 for the unbiased sample.

In each case, biasing the means caused the S_k^2 curve to reach a minimum at a smaller value of k^* than for the unbiased test; the stronger the bias, the faster the curve rose as k increased. Contrariwise, biasing the variance caused the curves to descend more and more steeply, so that for the strongest multiplicative bias, test 8, k^* is 49 or 50 years. The three independent random samples behave so similarly that these conclusions seem reasonably sound.

Examination of the curves of the previous investigators, in the light of the conclusions from these Monte Carlo tests, suggests that most of the climatic records heretofore investigated contain slight shifts in means. These may be true climatic changes, or they may be the result of changes in instruments, exposures, or observational practice. Whatever their nature, they produce climatic records in which, in general, the minimum variance estimate of next year's value is a mean over the most recent $k^* = 20$ or so years, rather than for a longer period. These findings are generally corroborated by further analysis, according to Eq. (2), of several long series of climatic data from various parts of the world, to be presented in a forthcoming report.

ACKNOWLEDGEMENTS

Most of the work covered in this report represents the efforts of William F. Slusser, graduate student in Geography. At the start of the project he combed through the climatologic literature (including the entire 16 years of Meteorological Abstracts and Bibliography) in search of any previous work. He directed the small group of student assistants who tediously scaled and replotted the results of previous studies, and who plotted the data from the three Monte Carlo samples. Together with Paul Roy, a mathematics student, he wrote and revised computer programs for the Monte Carlo samples, as well as for the many additional climatic samples to be presented in a later report. Typing was done by Luise A. Graff, Geography senior.

Valuable advice on how to present the results of this study came from the audiences at two preliminary presentations of the data, in February, 1966, at the Air Force Cambridge Research Laboratories and at the University of Oklahoma.

REFERENCES (Chronological)

1945 - N(ellie) Carruthers: "The Optimum Period for a British Rainfall Normal." Quart. J. Roy. Met. Soc. 71:144-150.

1954 - Robert W. Lenhard and Werner A. Baum: "Some Considerations on Normal Monthly Temperatures." J. Met. 11:392-398.

1954 - James C. Coffin: "A Method of Estimating Precipitation Normals for Short-Record Stations." Trans., Am. Geophys. Un. 35:405-412.

1955 - Glenn W. Brier: "Statistical Methods in Meteorology" (abstract). J. Am. Stat. Assn. 50:569.

1957 - Robert T. Beaumont: "A Criterion for Selection of Length of Record for a Moving Arithmetic Mean for Hydrologic Data." Trans. Am. Geophys. Un. 38:198-200. Disc. by T. E. A. von Hylckama, Reply by Beaumont, 39: 337, 1958.

1959a- Isadore Enger: "Optimum Length of Record for Climatological Estimates of Temperature." J. Geophys. Res. 64:779-787.

1959b- Isadore Enger: "Singularities in Daily Temperatures." J. Met. 16:238-243.

1960 - J. M. Craddock and M. Grimmer: "The Estimation of Mean Annual Temperature from the Temperatures of Preceding Years." Weather 15:340-348.

1965 - O. A. Drozdov et al: " --- " (Problem of the optimum length of the period of averaging in climatological investigations.) Trudy Glavaia Geofiz. Obs. imeni A.I.Voeikova 181:14-45.

1966 - F. F. Davitaia: "Climatic Norms and the Interrelation of Certain Natural Factors." Meteorologiia i Gidrologiia 1966(3):3-9.

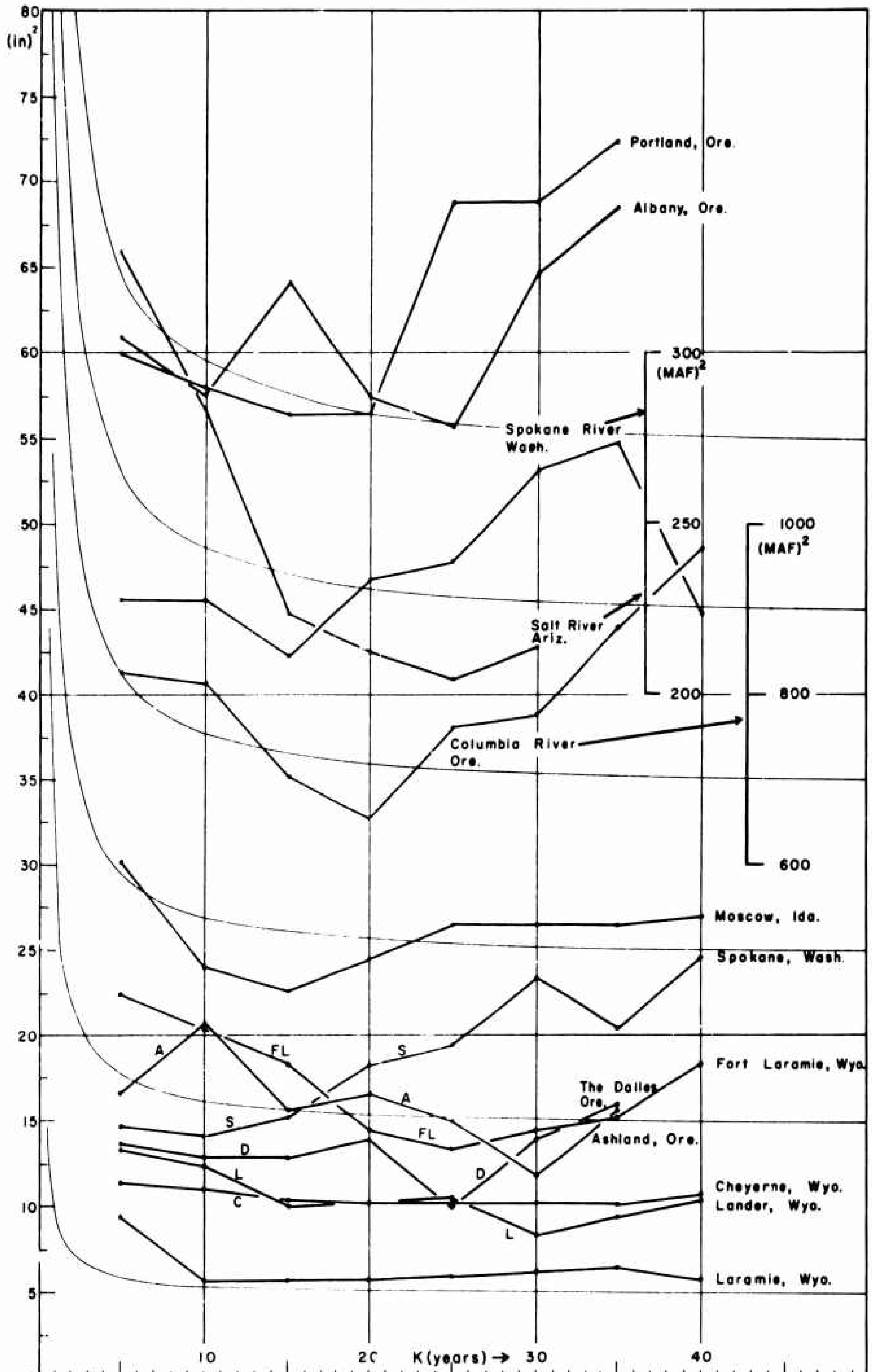


Fig. 1. Annual Streamflow and Precipitation (Beaumont 1957)

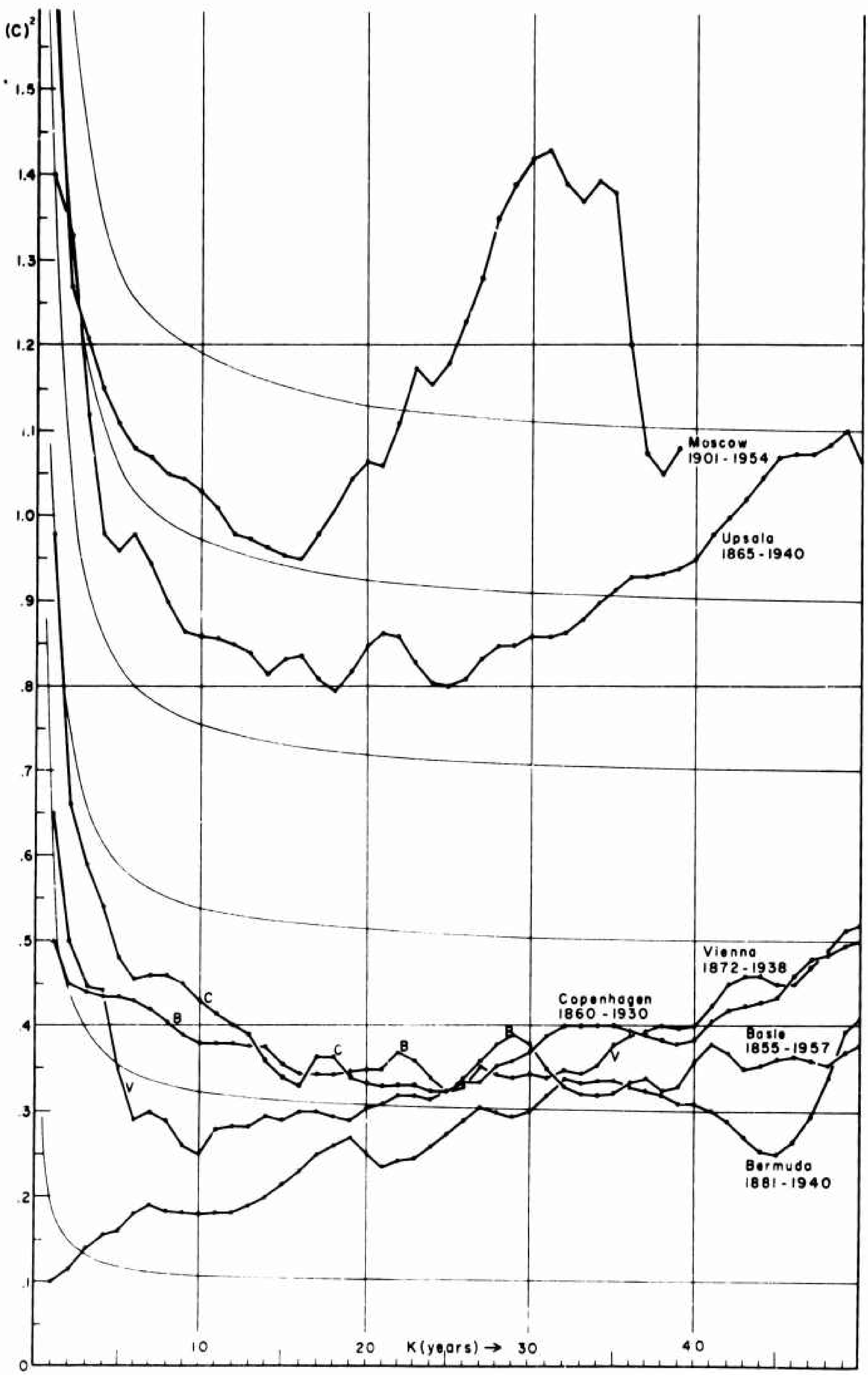
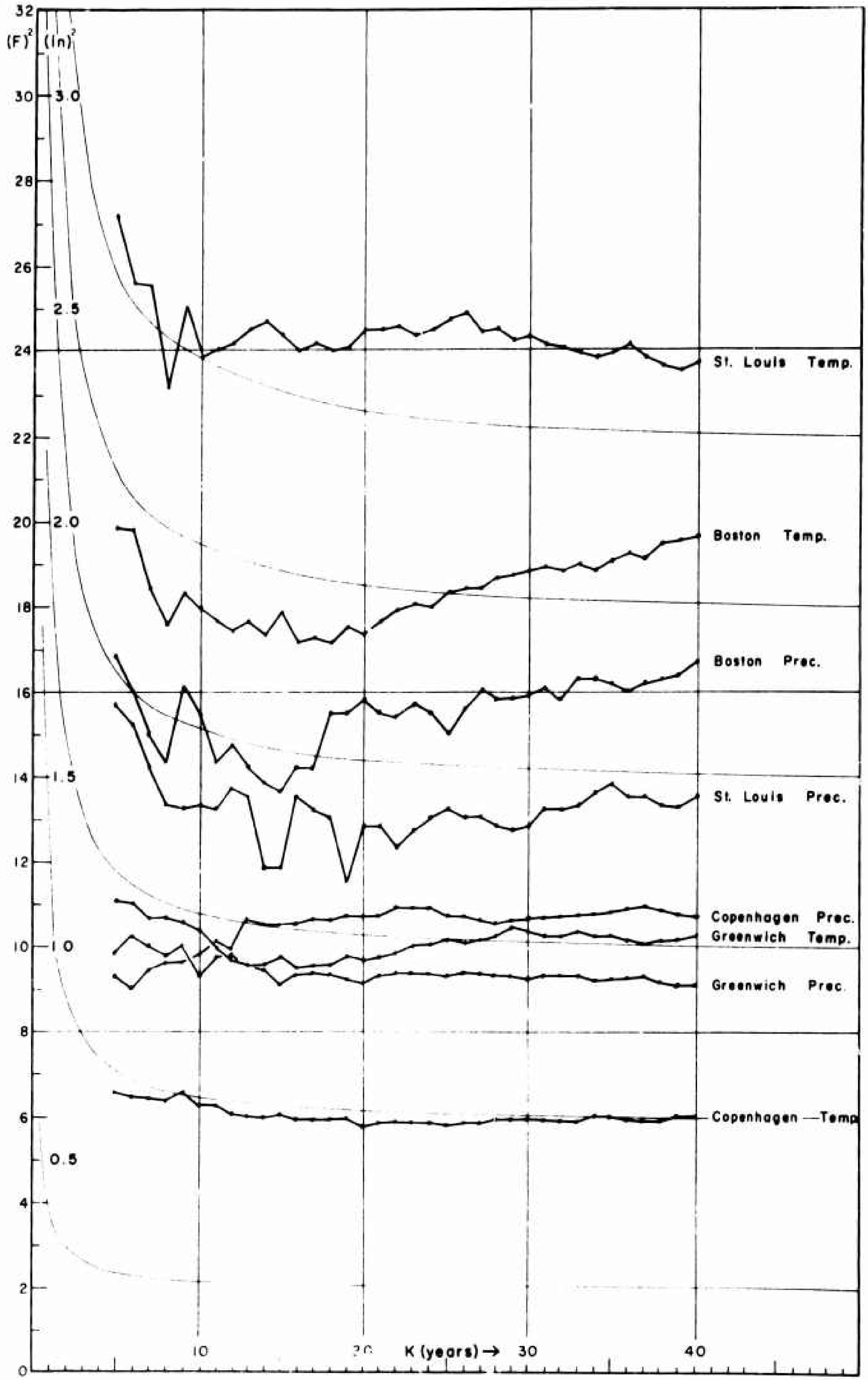


Fig. 2 Annual Temperature (Craddock & Grimmer 1960)



F[illegible] [illegible] January Temperature and Precipitation (Brier from Enger 1959a)

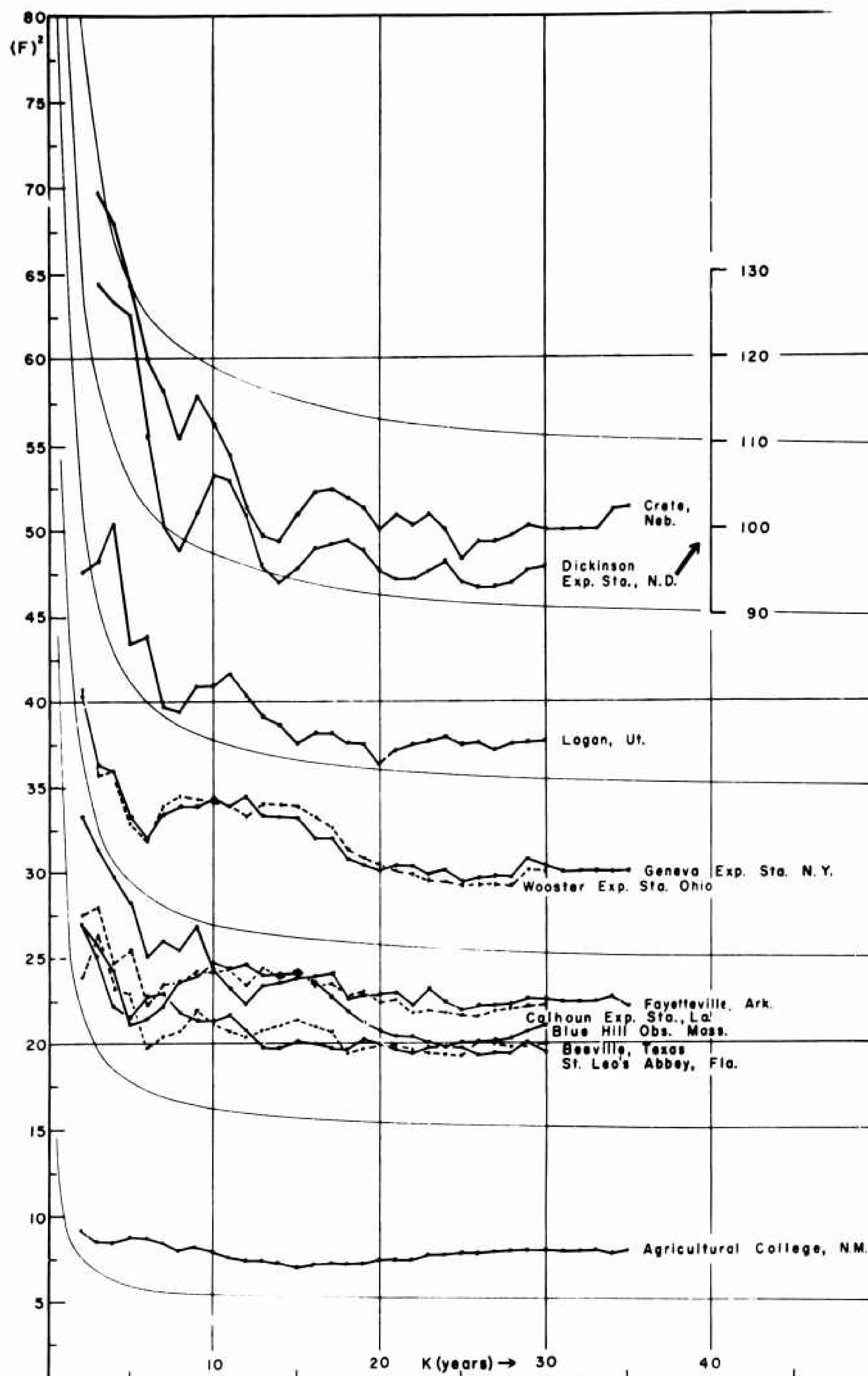


Fig. 4 January Temperature (Enger 1959a)

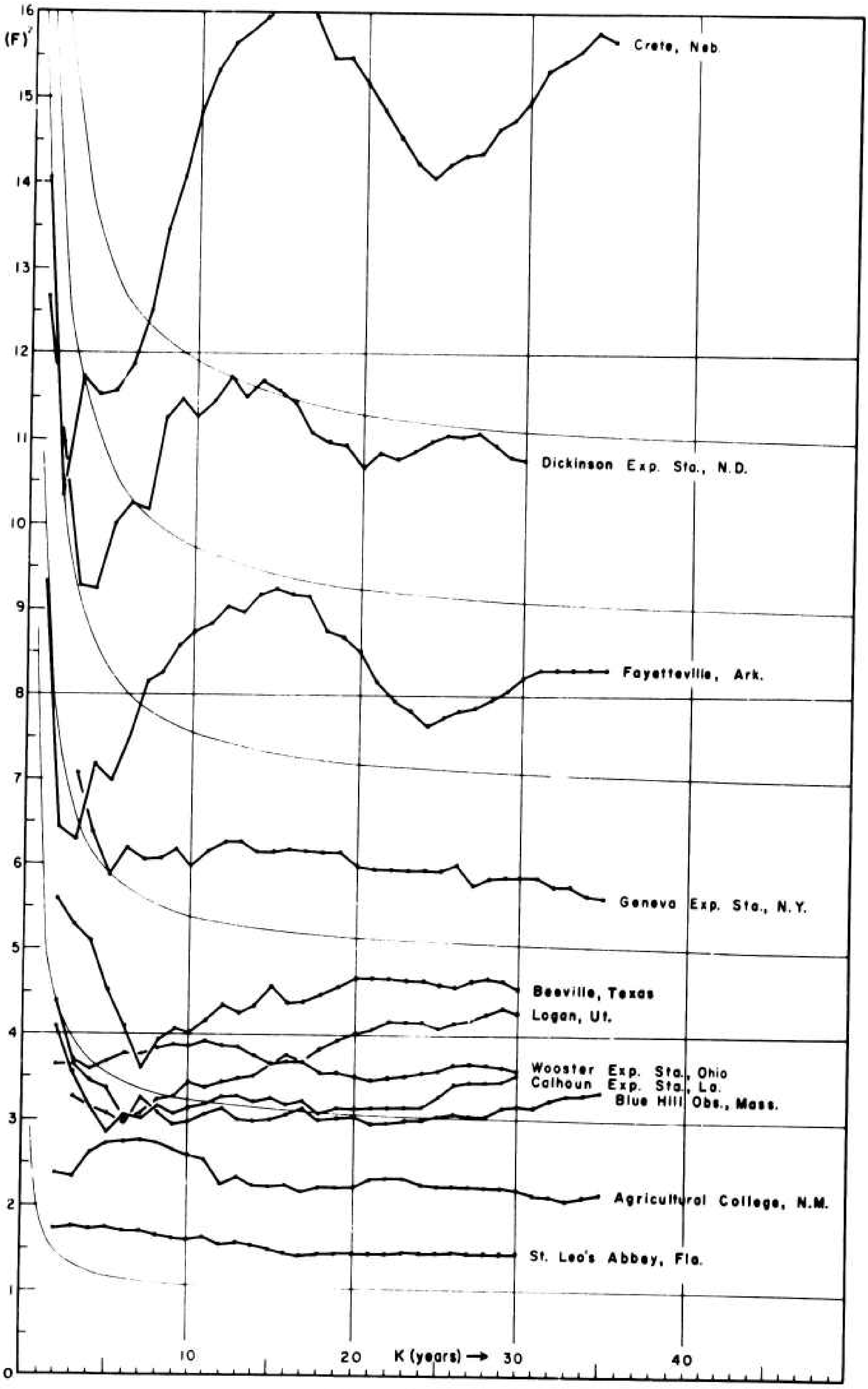


Fig. 5. July Temperature (Enger, 1959a)

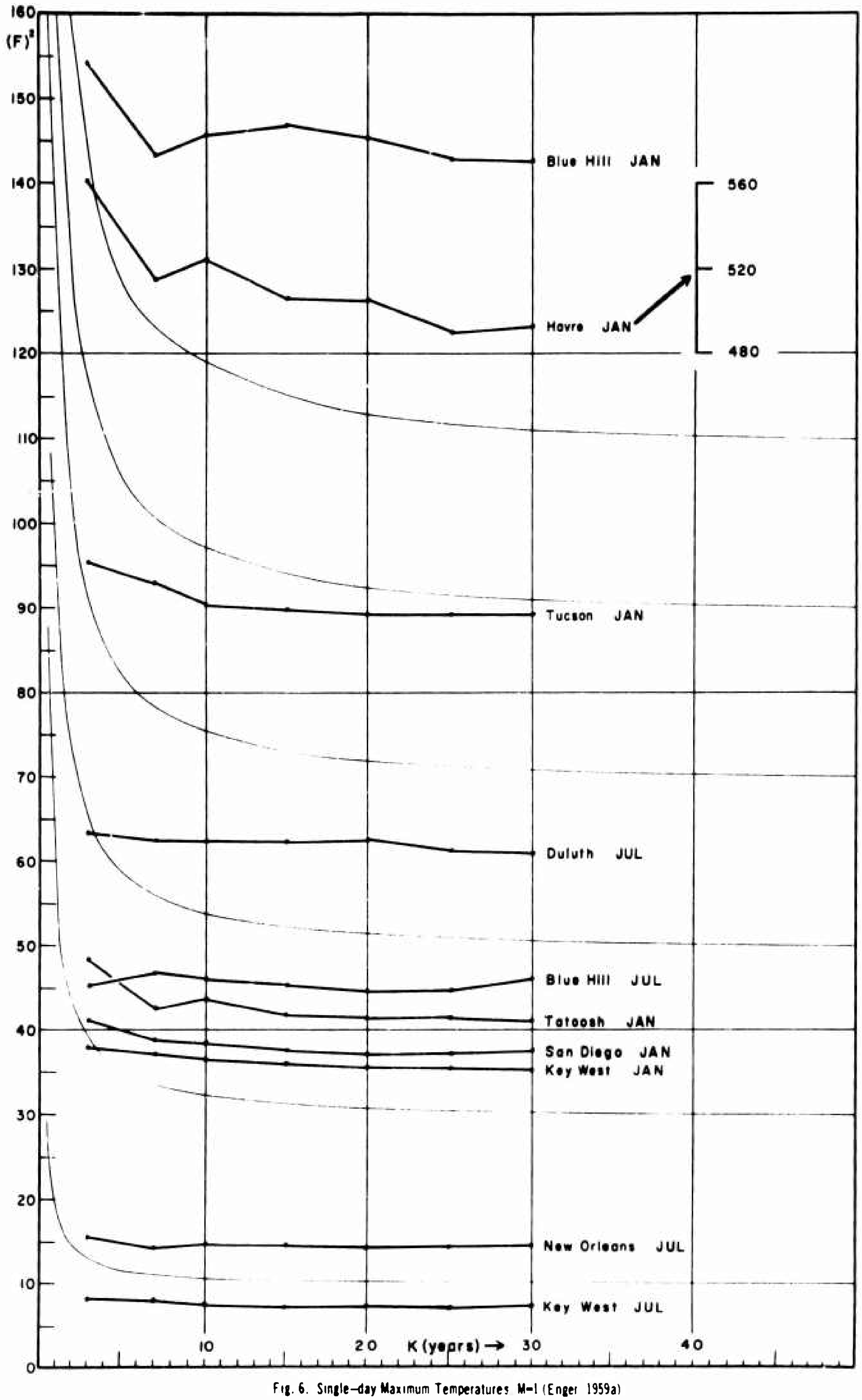


Fig. 6. Single-day Maximum Temperatures M-I (Enger 1959a)

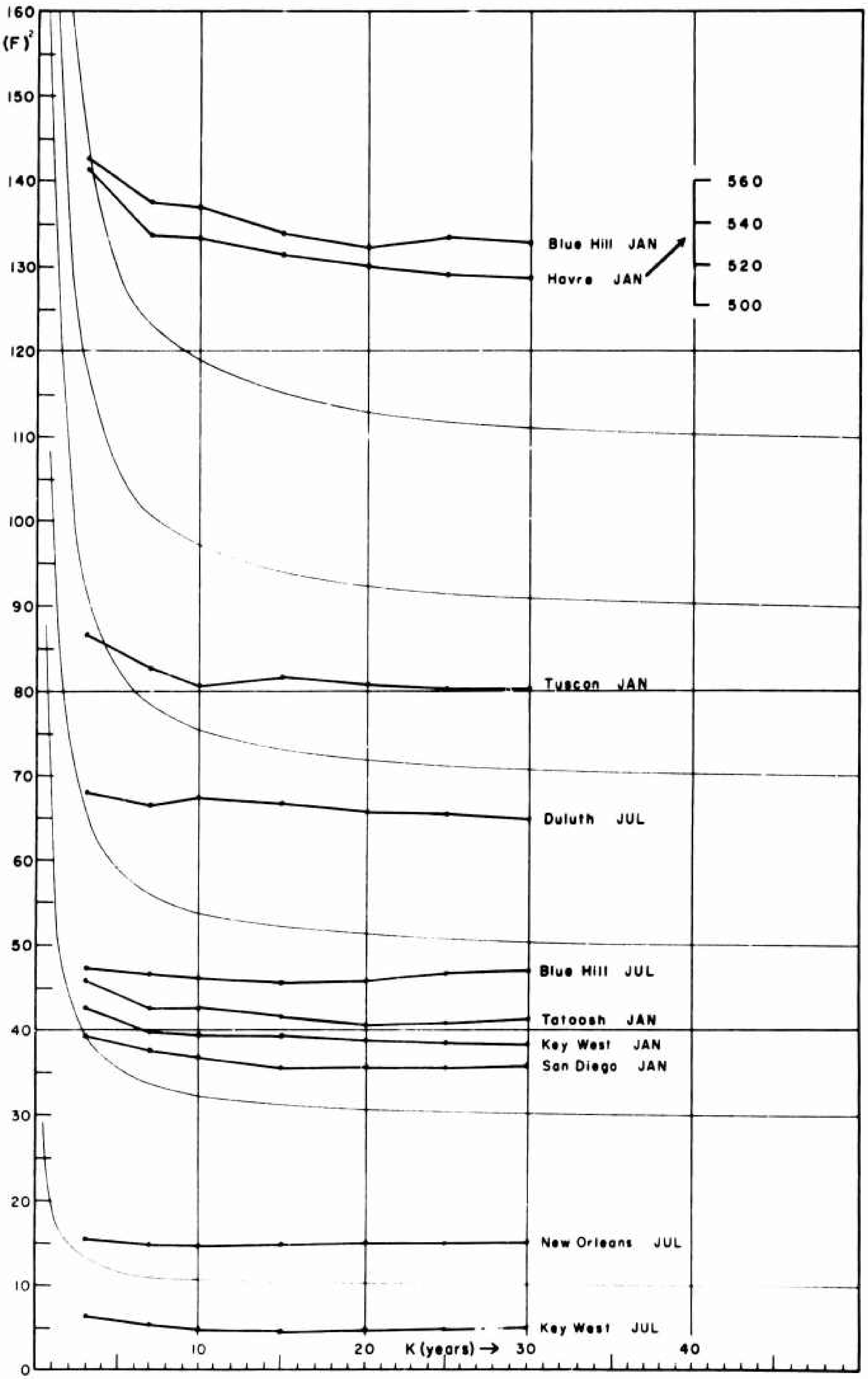


Fig. 7. Single-day Maximum Temperatures M=1,5,9 (Enger 1959a)

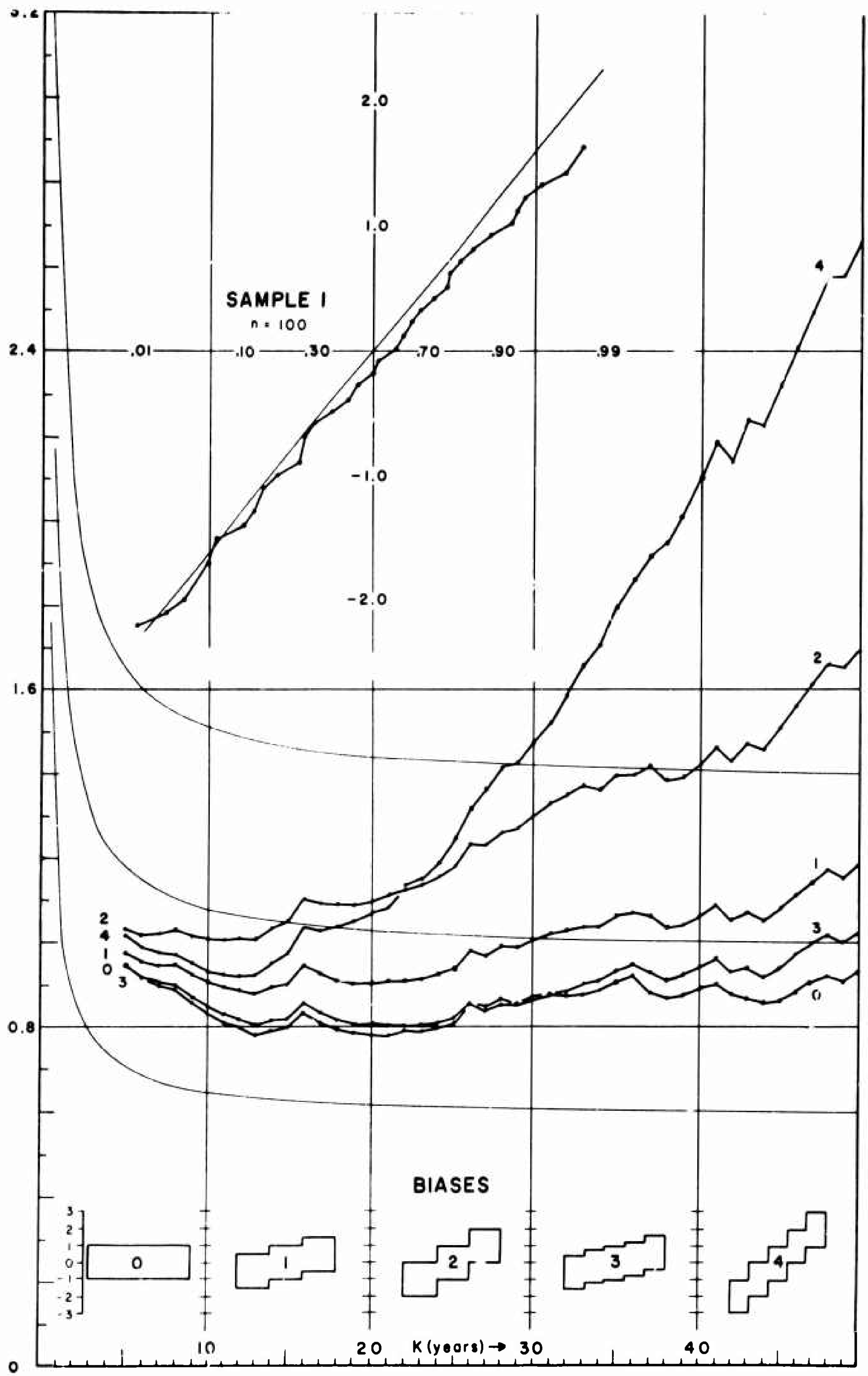


Fig. 8. Normal Sample No. 1, biased in mean

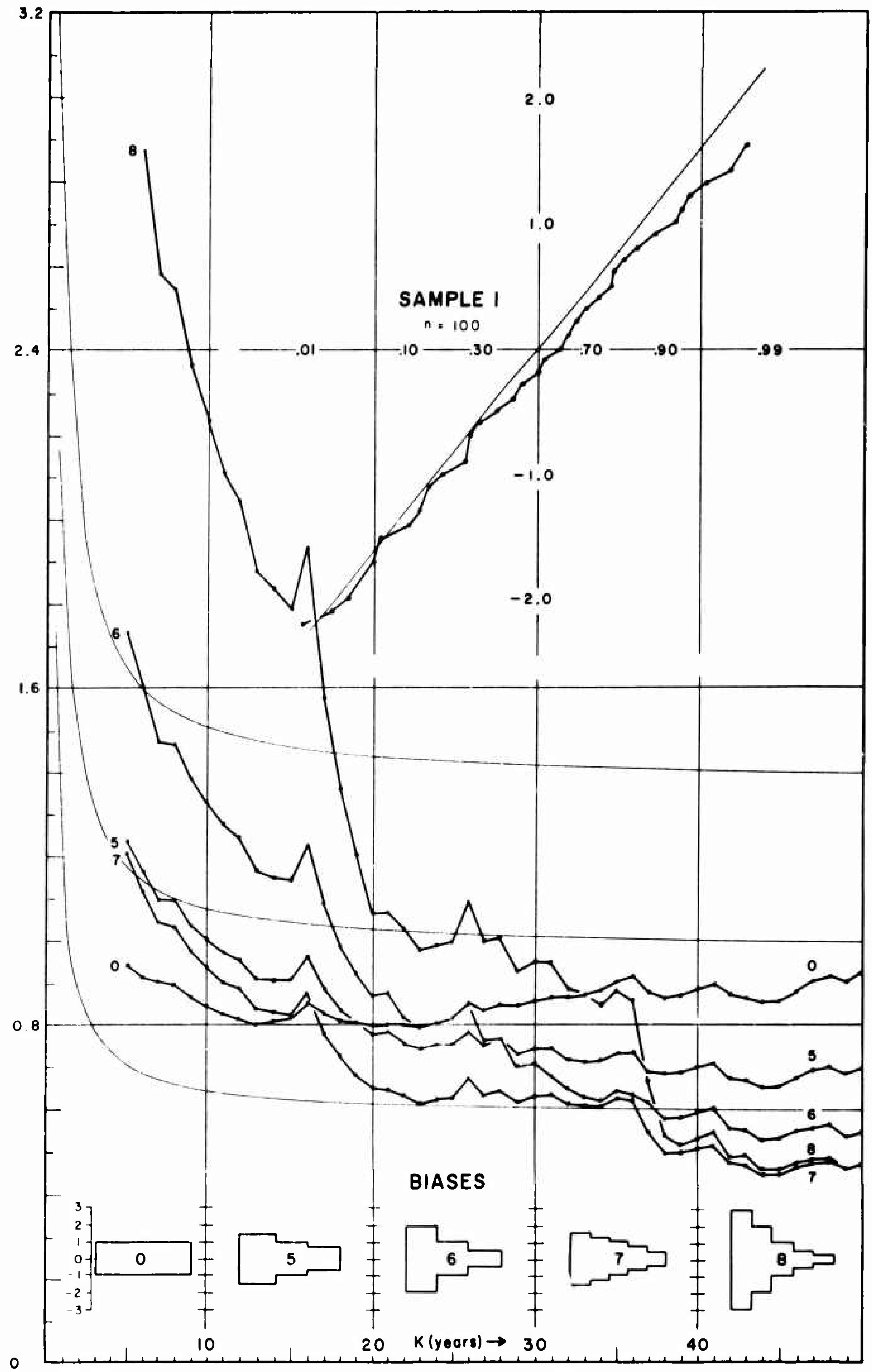


Fig. 9. Normal Sample No. 1, biased in variance

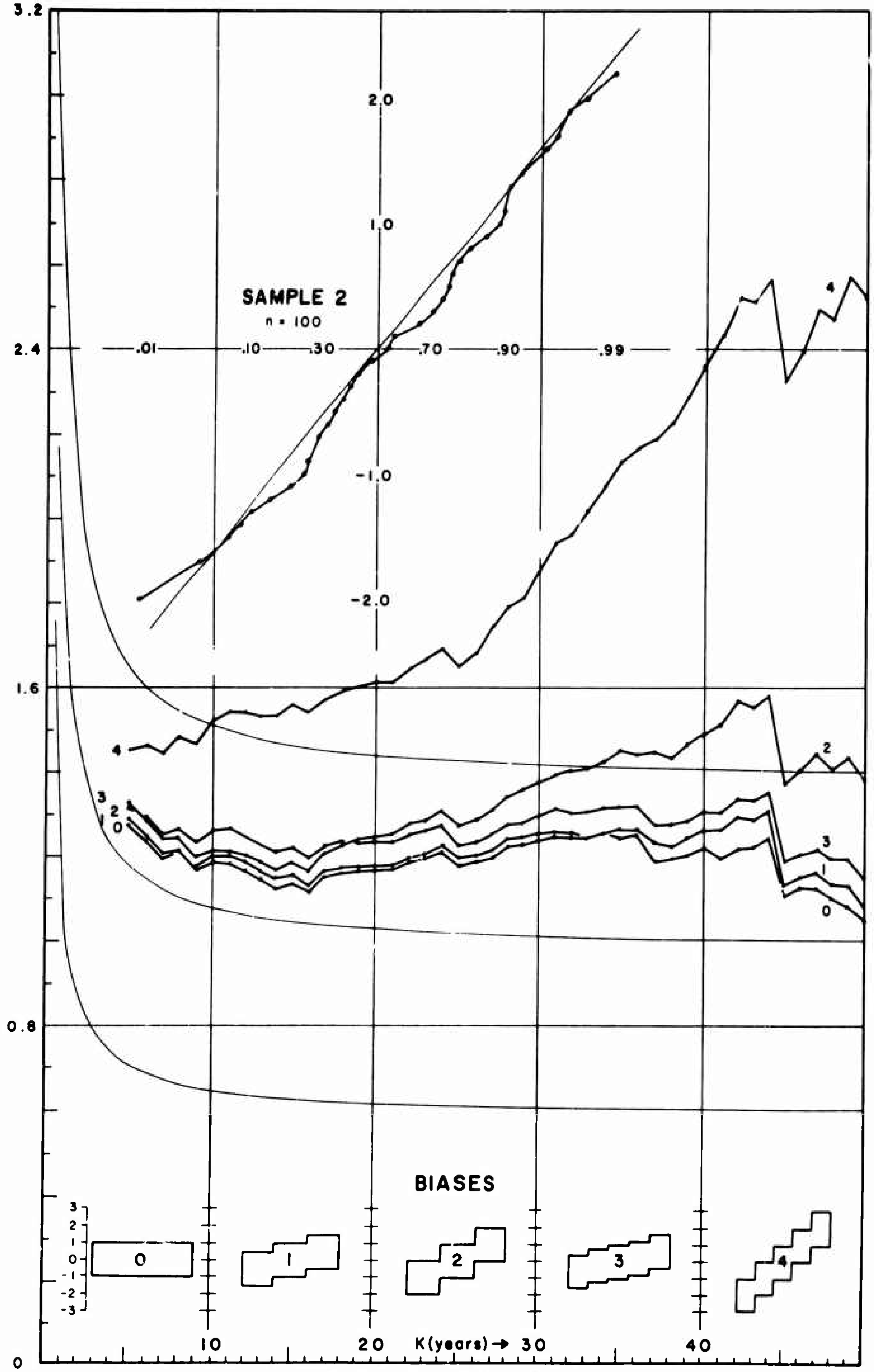


Fig. 10. Normal Sample No. 2, biased in mean

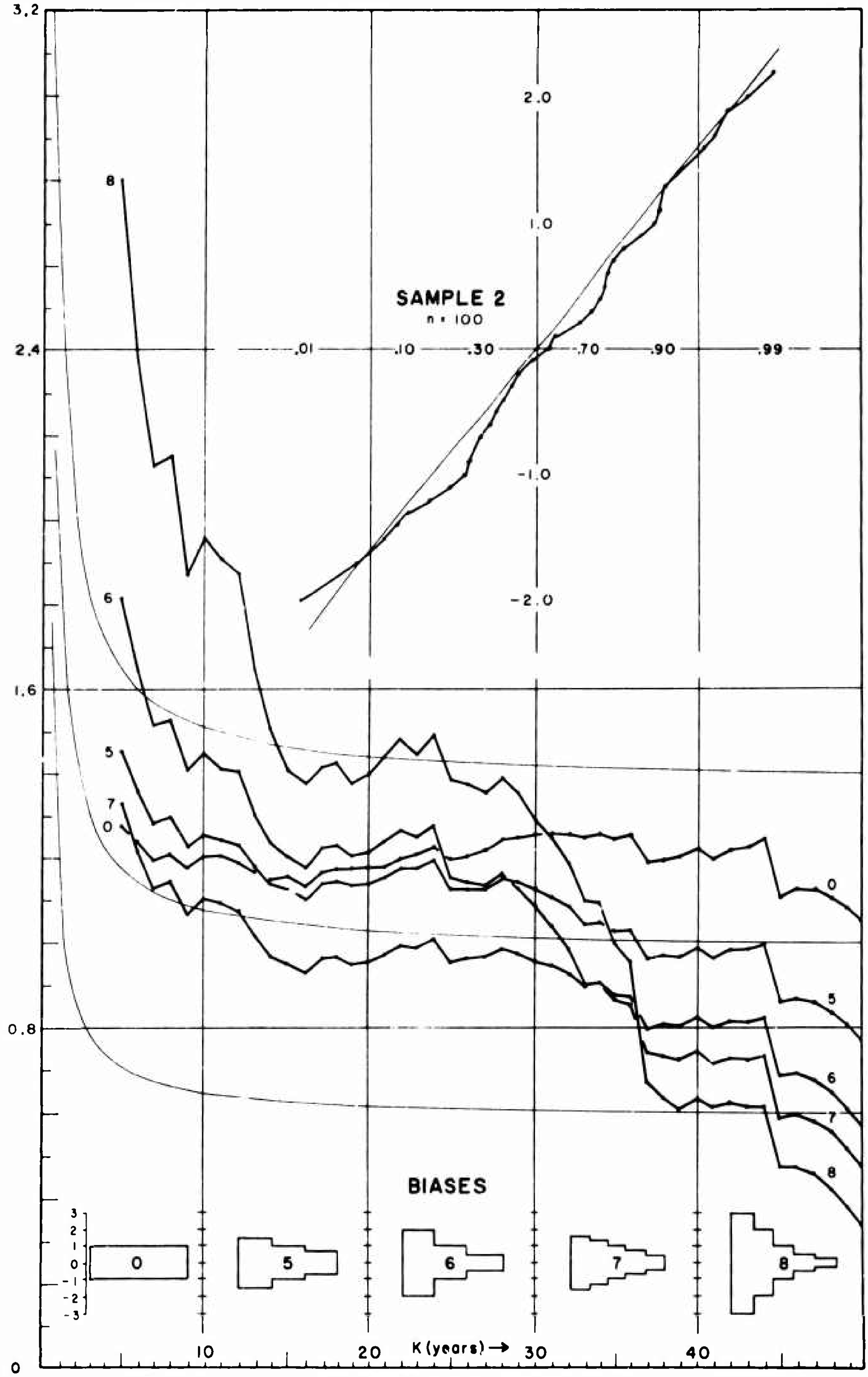


Fig. 11. Normal Sample No. 2, biased in variance

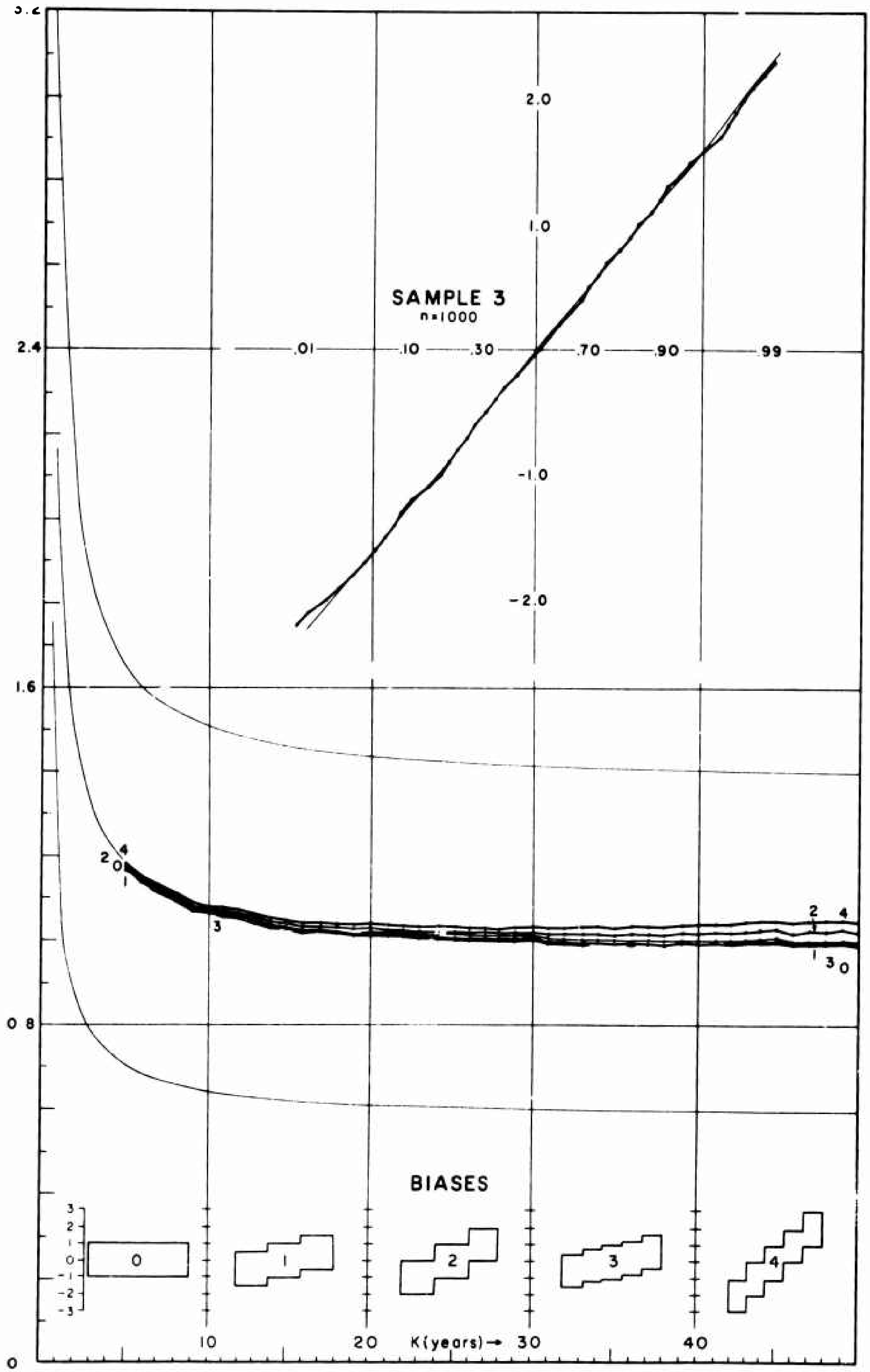


Fig. 12. Normal Sample No. 3 biased in mean

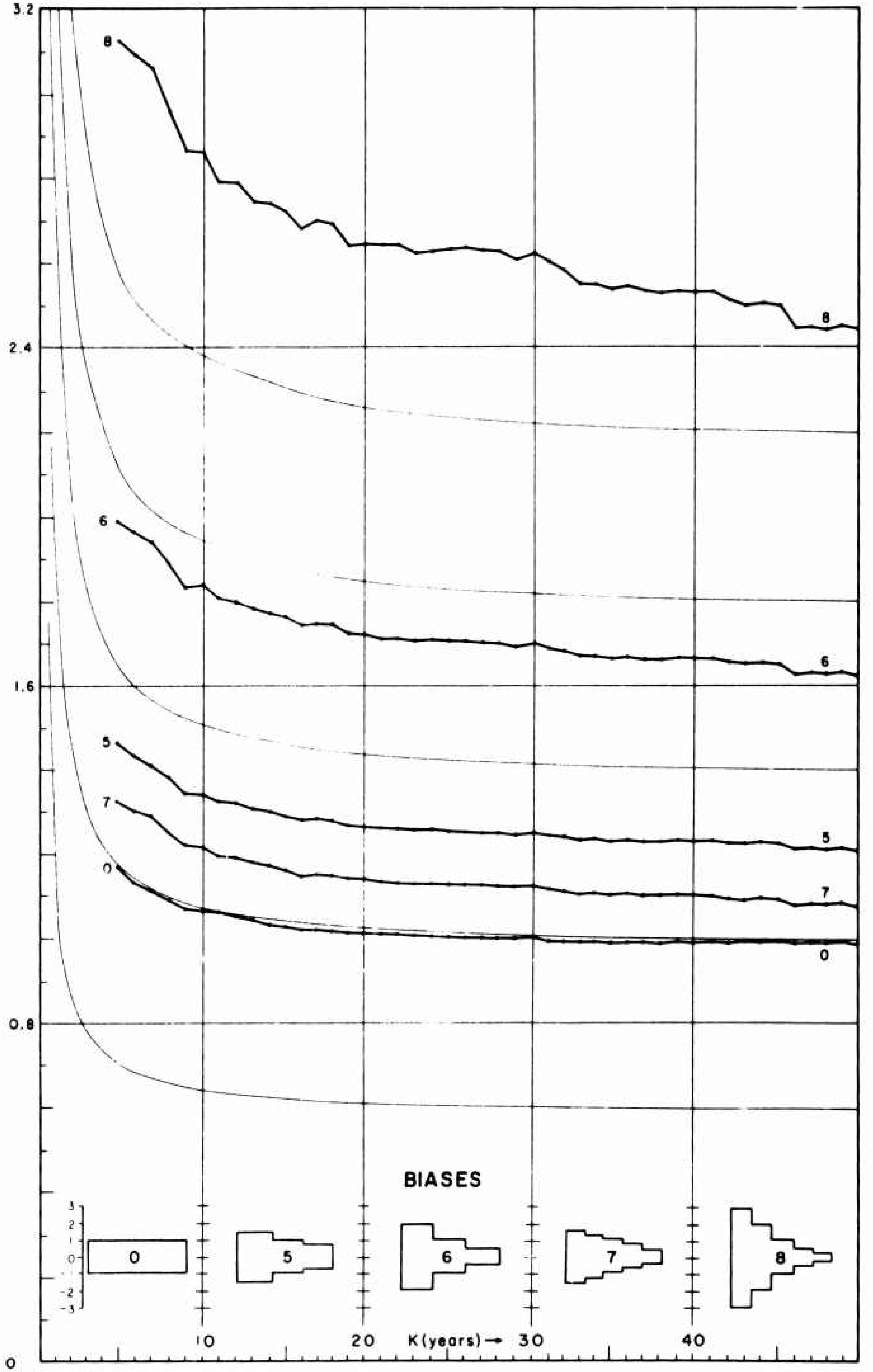


Fig. 13 Normal Sample No. 3 biased in variance

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) SAN FERNANDO VALLEY STATE COLLEGE FOUNDATION Northridge, California, 91324	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP —

3. REPORT TITLE

CLIMATIC NORMALS AS PREDICTORS. Part I: Background

4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Interim Scientific Report

5. AUTHOR(S) (Last name, first name, initial)

Court, Arnold

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1 May 1967	39	10

8a. CONTRACT OR GRANT NO. AF19(628)-5716 b. PROJECT NO. Task 8624-02 c. DOD Element No. 62405394 d. DOD Subelement No. 681000	9a. ORIGINATOR'S REPORT NUMBER(S) 82-67-1 Scientific Report No. 1 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFCRL 67-0313

10. AVAILABILITY/LIMITATION NOTICES

Qualified requestors may obtain copies of this report from DDC.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Hq AFCRL, (CRE) United States Air Force L.G. Hanscom Field Bedford, Mass., 01730

13. ABSTRACT

Preliminary to investigating the length of the climatic period whose average gives the best (minimum variance) estimate of the next year's value, previous studies are examined and the results of five are replotted onto a standard scale. All indicate that prediction one year ahead from an average based on only 20 years, or so, is better than one from a standard "climatic normal" of 30 years. Monte Carlo simulation of the prediction process suggests that slight changes with time in the means, whether real or caused by instrumental or observational changes, in most climatic records reduce the record length for optimum prediction.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
NORMALS, CLIMATIC PREDICTION, CLIMATIC OPTIMUM RECORD LENGTH						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

_______________________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

_______________________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

_______________________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

AD657358

SAN FERNANDO VALLEY STATE COLLEGE
NORTHRIDGE, CALIFORNIA

11 September 1968

ERRATA SHEET FOR

CLIMATIC NORMALS AS PREDICTORS

Part 1: Background AFCRL 67-0313

Part 3: Median vs. Mean AFCRL-68-0255

In Part 1, distributed in June 1967, the contract number on the cover should be corrected to read:

AF19 (628) - 5716

In Part 3, distributed in July 1968, in Form DD 1473 (the final page of report), block 9b should contain the number, which also appears on the cover:

AFCRL-68-0255

In Part 3 also, page 81, entitled "Climatic Prediction," should be marked as the start of

APPENDIX II

Part 2 and 4 are in preparation, and will receive the same distribution as the first two reports.

Arnold Court
Principal Investigator